INFORMATION AND CONTROL 12, 1–4 (1968)

Notes on a General Strong Converse[1]

J. WOLFOWITZ[2]

Department of Mathematics, Cornell University, Ithaca, New York 14850

1. INTRODUCTION

The two most widely applicable tools of information theory are, without doubt, Shannon's random coding theorem (in general form in Wolfowitz, 1964, Theorems 7.3.1 and 7.3.2), and Fano's weak converse (Wolfowitz, 1964, Theorems 7.4.1 and 7.4.2). In the present note we give what we hope will be a similar tool for proving strong converses. Its proof is also easy, in fact, almost trivial. It is a formalization and slight generalization of methods used by Kemperman, Yoshihara, and the author. Its justification lies in its applications. We give an application to the discrete memoryless channel, which will illustrate the technique of applying the general method to special problems. Application of the lemma below to semicontinuous and continuous channels, the channel of Section 9.2 of Wolfowitz (1964), and many, many others, is also easy. In the case of continuous channels one replaces probabilities by probability densities. The lemma below will easily be seen to hold with such a substitution.

2. THE STRONG CONVERSE

Let $G_1 = \{1, \cdots, g_1\}$, and $G_2 = \{1, \cdots, g_2\}$. Let $h(j \mid i)$ be a nonnegative function (c.p.f.[3]) defined for every pair (i, j) in $G = (G_1 \times G_2)$ and such that

$$\sum_{j=1}^{g_2} h(j \mid i) = 1, \qquad i = 1, \cdots, g_1 .$$

A code (N, λ) of length N and (maximum) probability of error λ,

[1] Research supported by the U. S. Air Force under Contract No. AF 18(600)-685, monitored by the Office of Scientific Research.

[2] Fellow of the John Simon Guggenheim Memorial Foundation.

[3] Channel probability function.

$0 \leqq \lambda \leqq 1$, is a system

$$\{(u_1, A_1), \cdots, (u_N, A_N)\} \tag{2.1}$$

where $u_1, \cdots, u_N$ are elements of G_1, $A_1, \cdots, A_N$ are disjoint subsets of G_2, and

$$\sum_{j \in A_i} h(j \mid u_i) \geqq 1 - \lambda, \qquad i = 1, \cdots, N. \tag{2.2}$$

A probability distribution d on G_1 (resp., on G_2) is a nonnegative function defined on G_1 (resp., on G_2), such that

$$\sum_{i \in G_1} d(i) = 1 \text{ (resp., } \sum_{i \in G_2} d(i) = 1).$$

Let D_1 (resp., D_2) denote the totality of probability distributions on G_1 (resp., on G_2). For each $i \in G_1$, $h(\cdot \mid i)$ is a member of D_2.

The symbol $P\{\ \}$ will denote the probability of the relation in braces and E will denote the expectation operator. When the relation in braces involves the chance variable X, the distribution of X, say d, will be indicated thus: $P\{\ \mid d\}$.

LEMMA. *Suppose that*

$$\min_{d \in D_2} \max_{i \in G_1} P\left\{\frac{h(X \mid i)}{d(X)} \geqq 2^{\theta} \,\middle|\, h(\cdot \mid i)\right\} < \gamma. \tag{2.3}$$

Then, if there exists a code (N, λ), *we must have*

$$N < (1 - \lambda - \gamma)^{-1} 2^{\theta}, \tag{2.4}$$

provided $\lambda + \gamma < 1$.

The proof is very simple. Let d^* be the element of D_2 which minimizes the expression in the left member of (2.3). Consider the code (2.1). For any integer $k \leqq N$ let

$$B_k = \left\{j \in A_k \,\middle|\, \frac{h(j \mid u_k)}{d^*(j)} < 2^{\theta}\right\}. \tag{2.5}$$

Then from (2.3) and (2.5) we have

$$2^{\theta} \cdot \sum_{j \in B_k} d^*(j) > \sum_{j \in B_k} h(j \mid u_k) > (1 - \lambda - \gamma). \tag{2.6}$$

Summing left and right members of (2.6) with respect to k we obtain that

$$2^{\theta} > N (1 - \lambda - \gamma),$$

which proves the lemma.

If, for a particular $d \in D_2$,

$$\max_{i \in G_1} P\left\{\frac{h(X \mid i)}{d(X)} \geqq 2^{\theta} \,\middle|\, h(\cdot \mid i)\right\} < \gamma, \tag{2.7}$$

the hypothesis (2.3) of the lemma is *a fortiori* satisfied.

3. THE DISCRETE MEMORYLESS CHANNEL

Let $A = \{1, \cdots, a\}$ and $B = \{1, \cdots, b\}$ be the input and output alphabets, respectively. Let $w(\cdot \mid \cdot)$ be the c.p.f. For this channel

$$G_1 = \underbrace{A \times A \times \cdots \times A}_{n \text{ times}}, \qquad G_2 = \underbrace{B \times B \times \cdots \times B}_{n \text{ times}}$$

and

$$h(v_0 \mid u_0) = \prod_{i=1}^{n} w(y_i \mid x_i),$$

when $u_0 = (x_1, \cdots, x_n)$, $v_0 = (y_1, \cdots, y_n)$. Let $H(\cdot)$ denote the entropy of a probability vector. Let $\pi = (\pi(1), \cdots, \pi(a))$ be a probability a-vector, and let $\pi' = (\pi'(1), \cdots, \pi'(b))$, where

$$\pi'(j) = \sum_{i=1}^{a} \pi(i) w(j \mid i).$$

Define

$$C = \max_{\pi} \left[H(\pi') - \sum_{i=1}^{a} \pi(i) H(w(\cdot \mid i)) \right]. \tag{3.1}$$

We now have to prove that, if there exists a code (N, λ), $\lambda < 1$, for the channel, then

$$N < \exp_2 \{nC + n^{1/2} K(\lambda)\}, \tag{3.2}$$

where $K(\lambda)$ is a function of λ only.

Let

$$v(u_0) = (Y_1, \cdots, Y_n)$$

be the chance sequence received when u_0 is sent. Let $\tilde{\pi}$ be a maximizing π in (3.1). A simple lemma due to Shannon (1956) (or Wolfowitz, 1964, Lemma 4.9.1) asserts that, for $i = 1, \cdots, a$,

$$E\left\{\log \frac{w(X \mid i)}{\tilde{\pi}'(X)} \,\middle|\, w(\cdot \mid i)\right\} \leqq C. \tag{3.3}$$

We therefore have

$$E \sum_{i=1}^{n} \log \frac{w(Y_i \mid x_i)}{\bar{\pi}'(Y_i)} \leqq nC. \tag{3.4}$$

Obviously, for some nonnegative constant t,

$$\text{variance} \left(\sum_{i=1}^{n} \log \frac{w(Y_i \mid x_i)}{\bar{\pi}'(Y_i)} \right) = nt^2. \tag{3.5}$$

Hence, for a suitable constant K and all n,

$$P\left\{ \left[\log h(v(u_0) \mid u_0) - \log \prod_{i=1}^{n} \bar{\pi}'(Y_i) \right] \geqq nC + n^{1/2}Kt \right\} < \frac{1-\lambda}{2}. \tag{3.6}$$

The desired result (3.2) follows at once from (2.7), (3.6), and the lemma.

RECEIVED: September 19, 1967

REFERENCES

WOLFOWITZ, J. (1964), "Coding Theorems of Information Theory." Second edition. Springer, New York.

SHANNON, C. E. (1956), On the zero-error capacity of a noisy channel. *IRE Trans.* PGIT, 8–19.